

Statistische und stochastische Grundlagen

VORLESUNGSMITSCHRIEB ZUM MODUL AN DER UNIVERSITÄT STUTTGART

INHALTSVERZEICHNIS

I	Einführung	3
1	Themenbereiche	4
1.1	Deskriptive Statistik	4
1.2	Explorative Statistik	4
1.3	Induktive Statistik	4
2	Grundbegriffe	5
2.1	Grundbegriffe der Statistik	5
2.2	Charakterisierung der Merkmale	5
2.3	Skalen von Merkmalen	5
2.4	Datengewinnung, Datenerhebung	5
3	Statistik	6
3.1	Erste Kenngrößen	6
3.1.1	Kumulierte Häufigkeiten	6
3.2	Klassifizierung	7
3.3	Lagemaße	7

1

Kapitel 1: Einführung

1: THEMENBEREICHE

1.1 Deskriptive Statistik

1.2 Explorative Statistik

1.3 Induktive Statistik

2: GRUNDBEGRIFFE

2.1 Grundbegriffe der Statistik

Statistische Einheit Objekte die erfasst werden

Grundgesamtheit Menge aller für die Fragestellung relevanten statistischen Einheiten

Teilgesamtheit Teilmenge der Grundgesamtheit

Stichprobe Tatsächlich untersuchte Teilmenge der Grundgesamtheit

Merkmal, Variable Größe von Interesse

Ausprägung, Wert Konkreter Wert des Merkmals für eine statistische Einheit

2.2 Charakterisierung der Merkmale

diskret abzählbar

stetig Werte aus einem Intervall

quasi-stetig stetig, aber nicht stetig messbar

2.3 Skalen von Merkmalen

Nominalskala Namen, Kategorien

Ordinalskala Ausprägungen mit Ordnung, aber Abstände nicht interpretierbar

Kardinalskala metrisch, messbar

2.4 Datengewinnung, Datenerhebung

Experiment - Erhebung

3: STATISTIK

3.1 Erste Kenngrößen

Als *Urliste* bezeichnet man die Menge der Merkmale X der Untersuchungseinheiten $U = \{x_1, \dots, x_n\}$. Die *auf tretenden Ausprägungen* von X sind die Werte $\{a_1, \dots, a_k\} \subseteq \{x_1, \dots, x_n\}, k \leq n$. Oftmals treten in einem großen Datensatz der Größe n nicht auch n verschiedene Werte x_i auf. Damit definieren sich

Definition 3.1: Absolute Häufigkeit

Die absolute Häufigkeit einer auftretenden Ausprägung a in einer Urliste U ist

$$h(a) = |\{x \in U \mid x = a\}|.$$

Es gilt immer, dass die Summe aller absoluten Häufigkeiten gleich der Datensatzgröße ist

$$\sum_{i=1}^n h(a_i) = |U|.$$

Die absolute Häufigkeitsverteilung ist dargestellt durch die Folge von Werten

$$h_1, \dots, h_k = h(a_1), \dots, h(a_k)$$

Eine grafische Darstellung der absoluten Häufigkeitsverteilung nennt man ein *Histogramm*.

Definition 3.2: Relative Häufigkeit

Die relative Häufigkeit einer auftretenden Ausprägung a in einer Urliste U ist

$$f(a) = \frac{h(a)}{|U|}.$$

Es gilt ähnlich wie bei der absoluten Häufigkeit für die Summe

$$\sum_{i=1}^n f(a_i) = 1.$$

3.1.1 Kumulierte Häufigkeiten

Die absolut kumulierte Häufigkeitsverteilung ist die Funktion

$$H(x) = \sum_{i: a_i \leq x} h_i$$

Ebenso, die relative kumulierte Häufigkeitsverteilung

$$F(x) = \sum_{i: a_i \leq x} f_i$$

3.2 Klassifizierung

Sind alle auftretenden Ausprägungen Elemente eines Intervalls $[a, b]$, lässt sich dieses in gleich große Klassen der Größe d unterteilen.

Eine Klassifizierung ist allgemein

$$[a, c_1), \dots, [c_i, c_{i+1}), [c_{i+1}, c_{i+2}), \dots \quad \forall i : c_{i+1} - c_i = d$$

Klassifizierte Daten sind i.A. einfacher zu interpretieren als große Mengen von Daten, die sich nur wenig voneinander unterscheiden.

3.3 Lagemaße